A Context-Sensitive Cloud-Based Data Analytic Mobile Alert and Optimal Route Discovery System for Rural and Urban ITS Penetration

Victor Balogun[1], Oluwafemi A. Sarumi[2(✉)], and Olumide O. Obe[2]

[1] University of Winnipeg, Winnipeg, MB, Canada
vi.balogun@uwinnipeg.ca

[2] The Federal University of Technology, Akure, Nigeria
{oasarumi,ooobe}@futa.edu.ng

Abstract. The rapid growth in the number of road users and poor road management have been deemed responsible for the upsurge in road congestions and fatalities in recent times. Many of the lives lost was due to inadequate or inefficient public-accessible alerts system and rerouting mechanisms during emergencies. The Intelligent Transportation System (ITS) was anticipated as a solution to the numerous road networks usage problems. Recently, some developed countries have implemented some forms of ITS initiatives. But the transition of the road networks to a fully integrated ITS has been slow and daunting due to the huge cost of implementation. The use of mobile devices as backbone infrastructure for ITS networks during public emergencies has been proposed. Despite the advantage of being a cheap alternative, low computing power of mobile devices limit their potentials to support the expected Big Data ITS traffic. In this paper, we propose a cloud-based context-sensitive ITS infrastructure that uses the cloud as a primary aggregator of traffic messages plus a hybrid Data Analytics algorithm. The algorithm combines the enhanced features of Apache-Spark and Kafka frameworks blended with collaborative filtering using the ensemble machine learning classifier. The novelty of our approach stems from its ability to provide load balancing routing services based on the users' profiles, and avoid congestion-using the Dynamic Round Robin scheduling algorithm to reroute users with similar profiles.

Keywords: Context-sensitive · ITS · Mobile alert · Road incidences · Cloud · Data analytics

1 Introduction

The report in [1] stated that about 1.2 million people around the world are killed while about 50 million people are injured every year as a result of traffic related accidents. The World Health Organization (WHO) reported that due to

A. L. Martins et al. (Eds.): INTSYS 2020, LNICST 364, pp. 35–51, 2021.
https://doi.org/10.1007/978-3-030-71454-3_3

this massive loss of lives, the world economy declines by up to $500 billion every year [2]. Many of these lives that are lost within the road networks have been attributed to inadequate or inefficient traffic management systems to alert and provide alternative routes to users when there are traffic incidents and emergencies. The Intelligent Transportation System (ITS) was anticipated as a solution to the numerous problems that are associated with the use of the road networks. ITS has been defined by the U.S. Department of Transportation (US DOT) as the integration of advanced communication technologies into the transportation technologies and vehicles, including a broad range of wireless and wireline communications-based information and electronics technologies; and all transportation modes, from pedestrian activities to freight movement [3]. The report by US DOT in [4] also reiterated that the mobility and accessibility of a region can be enhanced when ITS technologies are efficiently implemented thereby helping road users to go to wherever they want and whenever they want, in a safe and reliable manner.

In the past few years, many developed countries of the world including Japan, South Korea, Singapore and the United States have taken the lead to embark on some forms of ITS initiatives. Japan being the world leader in the deployment of ITS began with the use of the Vehicle Information and Communication System (VICS) in 1996, and the use of probes to capture real-time information in 2003 [5]. ITS initiatives started in the United States after the development of the Electronic Route Guidance System (ERGS) in 1970. The ERGS was meant to provide road users with route guidance information via on-board and roadside equipment [6]. Singapore installed real-time bus arrival panels in January 2008 at bus stops to make public transportation a more attractive option for commuters [5]. As a result of the need for ITS, South Korea built legal and institutional support for ITS by creating a master plan in 1997 which led to the development of an ITS technical architecture and a regional and supra-regional implementation plan [7]. Despite the progress recorded in recent times as a result of all these initiatives, the transition of the road networks to a fully integrated ITS network have been slow and seemed to be a daunting endeavor. One of the major setbacks to ITS penetrations in our rural and urban communities can be traced to the huge cost of implementation, especially in terms of infrastructural procurement, installations and management. The need for accurate and real-time traffic information is critical to successful ITS deployment. ITS data comes from heterogeneous sources and are in different formats. Big Data analytic has been advantageous in predicting traffic flow in ITS [8]. Despite the potentials of the analysis of ITS Big data for pattern detection, dimension reduction and complex predictions, the heterogeneous nature of the ITS data still constitutes a major challenge to the integration of ITS data analytic system [9].

2 Challenges of Big Data Analytic in Intelligent Transportation System

ITS applications have been described by [10] as complex and data-intensive applications that exhibit the 5Vs of Big Data in terms of volume, variety, velocity, veracity, and value. Due to the unforeseen explosion in the amount of heterogeneous data generated by the ITS in the range of several Petabytes (PB) of data, the conventional data analytic approaches have been discovered to be inefficient in handling the complexities involved [11]. Therefore, there is a need for data analytic systems that can transform the heterogeneous and complex ITS data from a conventional technology-driven system to a complex data-driven system.

A major challenge of using Big Data analytic in ITS is determining how data is collected within an ITS infrastructure [12]. ITS users including vehicles and pedestrians are in regular motion within the road networks making the traffic data collected to be incomplete or inaccurate. As a result, such data can not be a reliable source for implementing an efficient Big Data analytic system. Therefore, there is a need to develop data collection systems that are automatic and with minimal human intervention so as to reduce data errors introduced by humans and to improve the quality of ITS data collected. A viable solution to the automatic data capturing need is to develop a user profiling system that can automatically capture traffic data from ITS users.

Another related challenge to using Big Data analytic in ITS is determining how to store the massive data that are captured within the ITS networks. According to [12], ITS data level has jumped from the range of Terabyte (TB) level to PB level thereby making the growth in data storage capacity to lag behind the ITS data growth. The implication of this is that the conventional approaches to data storage and database tools will no longer be adequate to handle the massive ITS data. Though several integrated Big Data capable, and multi-cloud storage and hybrid storage are emerging as solutions to Big Data storage, there is a need for smart management tools that can provide integrated analytic within storage devices. The authors in [13] stated that simply integrating and standardizing data systems does not remove the requirement for such systems to be capable of editing raw data so that it can yield useful results. In addition, there is a need to develop efficient data analytic algorithms that can leverage machine learning and predictive data analytic to support real-time data forecasting requirements of ITS applications.

The timeliness of processing ITS massive data to support Big Data applications is another major challenge. Heterogeneous and complex traffic data from diverse sources need to be compared with historical data and processed at an instant [14]. According to [8], several generic and dedicated Big Data analytic frameworks including Apache Storm, Apache Flink, Apache Spark and Kafka streams have been developed. Despite the successes of some of these frameworks, there is a need to investigate cloud-based context-sensitive hybrid Big Data analytic frameworks that can combine the advantages of existing frameworks to deliver faster real-time data processing for current and near future ITS networks.

3 Related Work

There have been several initiatives to provide traffic services to road users using mobile applications. The Waze Apps [15] currently operated by Google provide location-based traffic information services to travelers using their cell phones. The problem with this application is that traffic data are obtained from road users through their social media platform like Twitter. Such data are semi-structured and unreliable because the application does not enforce the user to activate their geolocation functionality so as to provide the actual location information of a specific traffic incident [16,17].

INRIX Traffic [18] is a next-generation navigation and traffic app that uses smartphones and vehicle GPS data to provide road users with real time traffic information. The App automatically profiles users to provide routing and other traffic related information especially as it concerns changing road conditions. Despite the automatic user profiling and sending of alerts provided by the App, there is no evidence of reduction in congestion when users are rerouted to a new route during road incidents. This is because a new route can drastically become congested as many road users within close proximity of the incident are rerouted to the same route.

Moovit App [19] is an example of mobile applications that provide road users with real-time public transportation journey planning by combining public transportation data with users' cell phones data. Gaode App [20] allows road users to plan their trips that involve the transition between different means of transportation including trains to buses and buses to cars or bicycles. There is no evidence of Big Data analytic of these mobile applications as next-generation mobile applications must be able to process the anticipated near future massive data that are generated by ITS users in order to support road users with real-time and reliable traffic information.

As a summary to issues with related work, most route discovery and routing approaches provide similar routing information to users on similar routes and such road users are directed to the same route within the ITS infrastructure. This can lead to congestion of a newly discovered alternative route and can even cause new incidents of accidents. Therefore, there is a need for an approach that reroutes users based on the particular situation of a user and that also considers load balancing of alternative routes. In addition, there is no evidence of Big Data analytic of these approaches for efficient processing of the near future ITS Big Data so as to support road users with real-time and reliable traffic information including road incident alerts. We therefore propose a context-sensitive mobile alert system that uses two-layer Aggregator design to provide road incident alerts to subscribers. In addition, the system uses an optimal route discovery and load balancing techniques to provide alternative routes to road users during traffic incidents. Users with similar profiles are distributed over discovered alternative routes based on the real-time traffic condition of each route, thereby reducing incidents of congestion and road accidents. The two-layer alerts/traffic data Aggregator also introduces redundancy by using the cloud as a primary Aggregator while still maintaining a secondary land-based Aggregator.

4 Context-Sensitive Cloud-Based Mobile Alert and Route Discovery System

4.1 General Description

The proposed mobile system allows the subscribers to undergo an initial registration process before using the system. This registration process acts as a security mechanism to ensure that users are authenticated before they can actively contribute to ITS data. During this process, users detailed data are collected so as to create an initial profile for each user. Next, the system through crowdsourcing prompts the user to take a survey so as to gather some vital information related to the user's traffic experiences and preferences. These information will be used to determine and refine the Traveling Model (TM) parameters. The TM parameters include factors that influence choice of route taken by different road users including pedestrians, drivers, and bikers. Specifically, the information collected will be used to assign weight to traffic parameters that will be used when calculating the cost and the subsequent rating of a route. After the registration process, authenticated users will be able to send alert and traffic messages including: road condition (e.g. wet, slippery), weather conditions (e.g. snowy, rainy), emergency incidences (e.g. accident, explosion, road close) to the system. These traffic alerts can be sent automatically, periodically or on-demand by the users' mobile devices.

Alert messages are classified based on type of alerts (e.g. road condition, weather condition, emergency incidents), location of alerts, and the level of severity. These alert messages are aggregated using the Cloud-Based Alerts Aggregator (CBAA) and analyzed by the Big Data analytic component of the system. Authenticated users can be provided with either a proactive service or a reactive service. The proactive service involves users automatically sending and receiving alerts based on their locations to incidence of emergencies and the system subsequently providing alternative routes based on our proposed optimal route discovery and rating algorithm. With the reactive service, which is an on-demand service is when authenticated users request for the shortest route to a Point of Interest (POI) at any moment during traffic incidents. Alerts are sent to subscribers based on the 5W of user profiling including Who, Where, When, What and Why. The Who profile involves identifying the current user, while the Where addresses the location of the subscriber. The When profile deals with temporal aspects of past, present and future i.e. the time of an incident, while the What profile deals with identifying activities of the user on object, for instance biking, driving, or walking. Lastly, the Why profile addresses the subtle content such as the user's need and emotion, e.g. a user in need of routing assistance. The system will profile the user before providing either the proactive or reactive services so as to accurately meet the specific need of the user.

4.2 The Architecture

The proposed Context-Sensitive Cloud-Based Mobile Alert and Route Discovery System architecture presented in Fig. 1 is an improvement over the architecture proposed in [21]. It is essentially the integration of the existing Mobile Alert System with Context-Sensitive sub component and a cloud-based Big Data analytic technology. This is to provide real-time traffic alerts and optimal route discovery services to ITS users during traffic incidents including accidents and natural disasters. The idea is that the cloud and Big Data-based Mobile Alert system will use the cloud as a primary alerts Aggregator where traffic data can be analyzed using a proposed Big Data analytic algorithm, and the result sent as alerts to users. In this architecture, we still maintain a secondary Land-Based Aggregator (LBA) that will contain real-time results of data analytic from the cloud system. The LBA serves as a redundant mechanism to ensure that the Mobile Alert system continues to operate in the presence of unanticipated faults and failures.

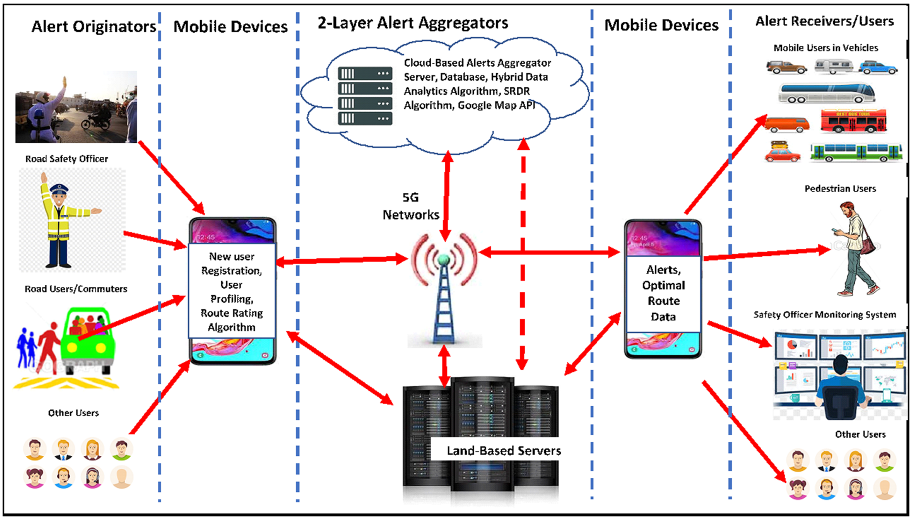

Fig. 1. The proposed architecture for context-sensitive cloud-based mobile alert and optimal route discovery ITS system

4.3 User Profiling (Context-Sensitive) Subsystem

The User Profiling subsystem illustrated in Fig. 2 uses the in-built mobile phones sensors to profile users so as to provide services that best meet the traffic needs of each user. The User Profiling subsystem consists of 10-tuple (decuple) attributes $(u_1, u_2, \ldots, u_{10})$ which are determined by four general travel model preferences including travelling preference, user location preference, weather condition and POI preference. We represented the travelling preference component with u_1 and u_2, where u1 captures the user's **travelling mode** (e.g. walking, riding, driving), and u_2 represents the **traffic condition** (congestion rate). The user location preference is set as u_3. This is an attribute which identifies the **user location at the moment** (the coordinate). We depicted the weather condition component with u_4 and u_5. Where u_4 is an attribute that captures the **weather condition of the user's current location** (e.g. snowy, rainy, clear, windy, cloudy), and u_5 is the attribute that describes the **weather condition of the user's destination**. We depicted the user's POI component with u_6 and u_7, where u_6 represents the user's destination and u_7 is an attribute that describes the user's perceived **fastest route to the POI**. We also consider some other factors that are necessary when road users need to make decisions about the choice of a route to take when navigating the road networks. We represented u_8 as the **time of the day** (e.g. morning – rush hour, afternoon, evening), u_9 is the **road condition to destination** (wet, slippery, bumpy), and u_{10} is the **road emergency/alert** (nature, location, severity). This profiling information is automatically captured by some certain sensors within the users' mobile devices and provided to other subsystems for further processing.

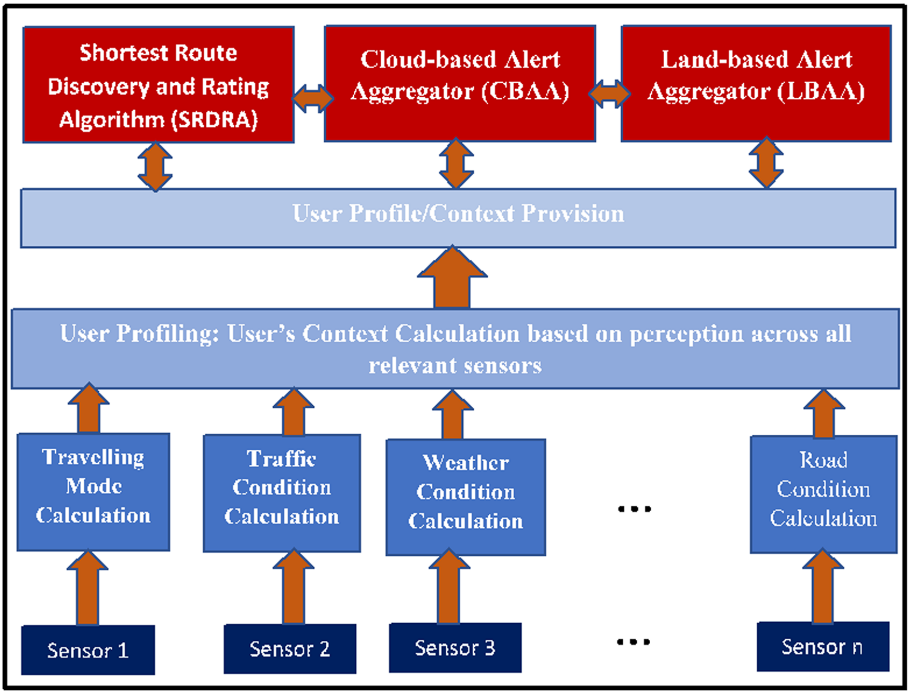

Fig. 2. User profiling (context-sensitive) subsystem

When routing information needs to be provided to a user, the user profile decuple, $(u_1, u_2, \ldots, u_{10})$ will be mapped with the travel model, quintuple, $(t_1, t_2, t_3, t_4, t_5)$ of a discovered route for rating purpose to determine how best the route correlates with the user profile. We established the parameters for the travel model by using t_1 to represent the **distance from user's current location to a POI** using the discovered route, t_2 represents the **traffic condition along the discovered route** (e.g. speed-limit, congestion), t_3 represents the **road condition along the discovered route** (dry, wet, bumpy, etc.), t_4 represents the **road incidents along the discovered route** (accident, roadblock, abduction, etc.), while t_5 represents the **weather condition along the discovered route** (rainy, snowy, cloudy, etc.)

4.4 The Route Rating Subsystem

The authors in [22] stated that the choice of route taken by a road user is a mix between decision-making under certainty and uncertainty. In [23], the choice of route was categorized based on attributes including characteristics of travelers, characteristics of routes (road, traffic, environment), characteristics of the trip and characteristics of other circumstances. In this paper, we propose a Route Rating subsystem that uses the combination of user profiling and crowd sourcing approaches to obtain personalized views of users traffic modes so as to develop a robust and user sensitive weighted averages for the traffic parameters that are used in our proposed route rating formula presented in Eq. (1).

4.5 Route Rating Formula

The results of user profiling discussed in Sect. 4.3 serve as input into the proposed Route Rating formula presented in Eq. (1) based on the Travel Model, quintuple, $(t_1, t_2, t_3, t_4, t_5)$.

The rating or cost of a route is derived mathematical as:

$$\textbf{Route Cost} = \left\{ \left(\frac{(t_1 w_1) + (t_2 w_2) + (t_3 w_3) + (t_4 w_4) + (t_5 w_5)}{max(totalCost)} \right) x10 \right\} \tag{1}$$

where:

$t_1, t_2, t_3, t_4, andt_5$ are the traffic parameters representing the Traffic Model

$w_1, w_2, \ldots, w_5$ are the weighted averages assigned to individual traffic parameters (these averages are derived from the result of crowd-sourcing during users' registration, the values will become more refined and reliable as the number of users increases).

max(total Cost) is the maximum of the sum of the product $t_1 w_1 \ldots\ldots w_5 t_5$. The route cost formula can be represented in Sigma notation as follows:

$$\mathbf{R}(\mathbf{TM}, \mathbf{UP}) = \left\{ \left(\frac{\sum_{i=1}^{5} f_i(P.W)}{max \sum_{i=1}^{5} f_i(P.W) \neq \emptyset} \right) x10 \right\} \tag{2}$$

Where:

R(TM, UP) is the route cost or rating derived from the mapping of the User Profile, UP and Travel Model, TM.

$f_i(PW)$ is a function representing the product of a traffic parameter, **P** and its corresponding weight parameter, **W**.

The weighted average assigned to each of the route cost parameters in Eq. (2) will differ depending on the travelling mode of the user. If the user is a pedestrian, for instance, the distance to destination and instances of road incidents might generate lower weighted averages compared to when the user is driving.

4.6 Rationale for the Route Rating Parameters

In a survey conducted in [24], the researchers discovered that the factors that are pertinent to the choice of route taken by bikers are the condition of the roads, volume of traffic, the speed of motor vehicles along the route, and the distance to destination. The results of the survey show that the speed of motor vehicles and volume of the traffic contributed largely to the overall decision made by bikers on the choice of route taken and were respectively rated 100% and 97%, with 100% rating indicating greatest contribution. Several researchers suggest that the choice of driving a car along a route is mostly affected by the travel time, road category, road safety, scenic quality, and the number of traffic lights and stop signs [25,26]. The results of the work in [22] indicate that the probability of selecting a route decreases with a rise in the travel time, and drivers tend to select the routes with lower tolls. The authors in [27] also identified seven factors that pedestrians consider favourable when selecting a route to take to a POI. These factors include shorter distance, lower travel time, even sidewalks, connected links, less crossing and barriers, low congestion level, and safety.

As a summary, we discovered from the various literature that the factors that are common to road users (bikers, drivers and pedestrians) when choosing a route to get to a POI include distance to the POI, road condition, level of congestion, and instances of traffic incidents. We therefore used these factors as bases for formulating our proposed Road Rating formula. To the best of our knowledge, we noticed that none of the literature considers the road weather conditions as a factor that can influence the choice of route by road users. Due to varying geographic factors, the weather condition along a particular route might be significantly different from that of another route. Since weather conditions affect the ease of biking, driving or walking along a route, we therefore consider the use of weather conditions as one of the factors to be considered while rating a route using our Route Rating formula as shown in Eqs. (1) and (2). We also restate here that the weights to be assigned to these traffic parameters including the weather conditions will be the weighted average derived as a result of the crowd-sourced data from users' initial or periodic survey discussed in Sect. 4.1.

5 The Context-Sensitive Mobile Alert Subsystem

The Context-Sensitive Mobile Alert subsystem of our proposed architecture is illustrated in Fig. 3. Traffic related data are automatically sensed by the Alert Sensing Layer and transferred to the Alert Detection Layer periodically. Traffic data can be received from the cell phone sensors, Google Map API, and related systems to profile the user. At the Alert Detection Layer, the received traffic data and the user behavior are analyzed to derive a profile for the user. The result of the user profiling is further analyzed and compared with normal traffic data to detect abnormal traffic conditions. This abnormal traffic data is sent to the Alert Processing Layer in the cloud where the abnormal traffic data is aggregated (**Message Aggregation**) using the Cloud-Based Alert Aggregator (CBAA). Before processing, the received message is authenticated to validate the source of the message (Message Authentication). This is a crucial component of the CBAA as the mechanism prevents instances of malicious users or hijackers from using the system to create public panics or as a tool for terrorism. The authenticated message is then processed using a proposed hybrid Data Analytic Engine that compares the current data with historical data and then translate or adapt the message into an alert standardized format (**Message Adaptation/Translation**). The translated alert message is then sent by the CBAA to each affected subscriber (within a determined radius of the incident, depending on the type and the severity of the incident) in a format that matches their profiles (**Alert Dissemination**). For instance, if the travelling mode is driving/biking, an audio alert is sent to the user but if the traveling mode is walking/running a textual alert is sent to the user.

5.1 The Cloud-Based Alerts Aggregator and the Optimal Route Algorithm

The results of user profiling (generated by the User Profiling subsystem) and the rating of the route currently occupied by the user (generated by Route Rating subsystem) are sent by a user's mobile device periodically (proactive service) or on-demand (reactive service) to the CBAA. As an improvement over existing cloud-based ITS systems, we anticipated instances of cloud system faults and failures by designing two-layer alerts/traffic data aggregators by implementing a CBAA while still maintaining a redundant Land-Based Alert Aggregator (LBAA). The LBAA acts as a cache for offline access, especially for rural ITS users to traffic messages when the CBAA experiences faults or failures.

The CBAA uses the user profile received to discover all available alternate routes from the current position of the user to the POI using the Dijkstra's Shortest Path Algorithm (DSPA) [28]. For a particular alternate route discovered by DSPA, the CBAA checks its database to see if there are current Travel Model data available. If such data exist, the CBAA uses the Route Rating formula described in Sect. 4.5 to calculate the route cost and rate the route. If there is no current Travel Model data, The CBAA will first use the Google Map API data to determine the real-time Travel Model data for the route. If there is

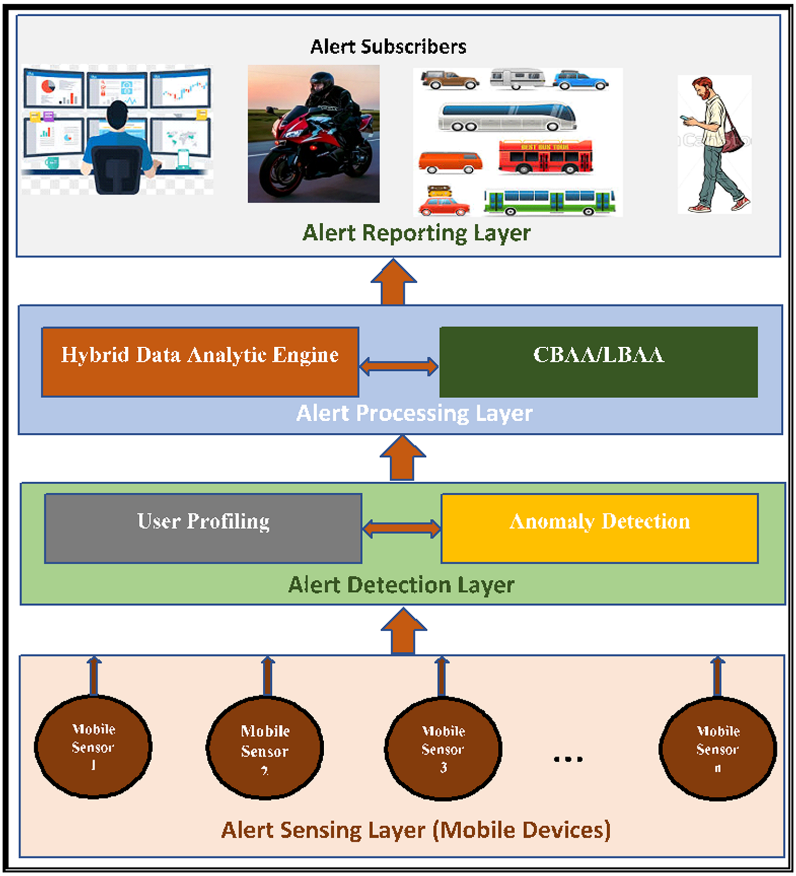

Fig. 3. Context-sensitive mobile alert subsystem

no current Travel Model data for the route, then the CBAA will generate the Travel Model data by triggering the hybrid Data Analytic Engine subsystem (see Fig. 4). The Data Analytic Engine subsystem, using our proposed hybrid Data Analytic algorithm, discussed in the next section, will analyze the massive historical traffic data stored in the cloud to generate the requested Travel Model data for the alternate route. The CBAA will perform the above process for all the discovered alternate routes and at the end will generate a list of available routes with their ratings in a descending order.

The CBAA then compares a user profile with the list of shortest routes generated to determine an optimal route for that user. The implication of this design is that two users traveling on a particular route might be rerouted to different routes depending on their individual profiles. The CBAA sends this optimal route information to a particular user in need of rerouting service via the user's mobile device. When the optimal route information gets to the user, the user will reroute to this newly discovered optimal route. Road users are rerouted to the newly discovered route using the Dynamic Round Robin (DRR) load balancing algorithm [29]. The DRR is a simple, resilient, weighted but dynamic scheduling algorithm that in our implementation, will allocate routes to users dynamically based on the real-time condition of the routes as rated by the Route Rating subsystem described in Sect. 4.3. This is an essential component of the proposed system as the load balancing functionality of the system prevents a newly discovered route from becoming overly congested due to poor scheduling, and it also reduces the chance of an accident occurring in the route.

5.2 The Cloud-Based Data Analytic Engine Subsystem

In order to efficiently process the complex and heterogeneous Big Data that are associated with ITS applications, we proposed a cloud-based Data Analytic Engine (DAE) that would provide real-time data analysis, on-demand decision support, and context-aware recommendations to the various users of the ITS infrastructure. To transform the heterogeneous and complex ITS data, our proposed cloud-based DAE as shown in Fig. 4 is an enhanced form of Apache Stark framework [30,31] hybridized with the Apache Kafka [32] for Big Data analytics. The cloud-based DAE is subsumed with data-driven techniques including features selection, collaborative filtering, and ensemble classifiers to provide intelligent and real-time traffic support to road users. In order to manage the storage and processing of the Big Data that are captured within the ITS network, the DAE leverages the scalable feature [33–35] of the Apache Spark framework that allows the system to expand by adding more nodes as the volume of the data increases. Additionally, to foster the timeliness of processing of the ITS Big Data, the DAE harnesses the fault-tolerant feature [36] of the Apache Spark framework fused with the distributed streaming platform of the Apache Kafka to provide swift and efficient real-time streaming and data analytic within the ITS infrastructure.

Some additional features of Apache Spark that we implement in the DAE for ITS Big Data analytic includes the use of a dedicated resource dispenser and a result accumulator called the driver program. The driver program enables a smooth coordination of all the data processing operations within the ITS and reconfiguration of lost partitions effortlessly without depleting the information. The driver program also ensures that there is no loss of critical information between the data analytic engine and the alert subscribers. The DAE with the aid of the driver program stores intermediate results in memory instead of disk, and supports ample system workloads such as interactive processes, batch processing, iterative procedures, machine learning, and graph processing. These are essential functions necessary for the efficient processing of the ITS Big Data. We introduced a filter-based feature selection model to the implementation of the DAE so as to eliminate noise and data redundancy from the complex and heterogeneous data aggregated from the alert originators. This is crucial for the optimal performance of our proposed collaborative filtering model. Collaborative

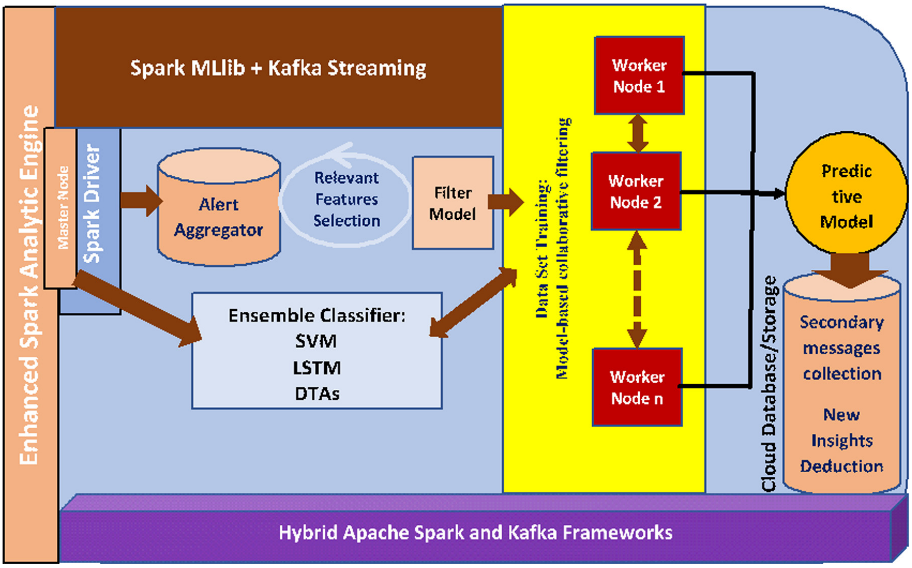

Fig. 4. Cloud-based data analytic engine subsystem

filtering [37] is a domain-independent prediction that explores historical and current data of the users to make intelligent recommendations for the users based on their current need.

In Fig. 5, we present our proposed Data Analytic algorithm that controls the DAE. The algorithm uses a filter-based features selection model to remove redundant and noisy data from the avalanche of dataset coming from the alert originators. We implement a model-based collaborative filtering technique so as to benefit from the advantages of ensemble machine learning classifiers [38]. This approach allows us to harness the strength of multiple machine learning techniques so as to perform predictive analytic and provide intelligent recommendations of the optimal route to users. Specifically, our proposed algorithm uses ensemble machine learning classifiers consisting of a hybrid of three machine learning methods including Support Vector Machine (SVM), Long Short-Term Memory (LSTM) and Decision Tree Algorithm (DTA). We are taking the advantage of the strength of SVM at being able to process heterogeneous and nonlinear data, and combining it with LSTM which is highly efficient in the analytics of historical data. Another advantage of this hybridization is the use of DTA which is capable of handling both regression and classification problems. Regression and classification considerations are desirable data analytic procedures for implementing robust ITS infrastructures.

```
function Cloud_Analytic_Aggregator ( ):
    Start DAE( )
    From DAE( )
  Do
      {
      Aggregate primary alert messages from the Originators,
      select relevant features from the datasets using filter-based model
      }
    For each worker node j on the Spark engine:
{
     Perform a model-based collaborative filtering (Ensemble model) on the dataset
     coordinated through the   master node.
    Aggregate the results of the training from the worker nodes on the master node
    (the context-aware predictive model)
}
Save the Predictive Model and relay it to the land-based servers

Collect   secondary messages from the mobile devices and store on the land-based
servers
Use the Predictive Model to deduce new insights from the secondary data
Provide intelligent context-aware decision support to the road users
return
```

Fig. 5. ITS data analytic algorithm

6 Conclusion and Future Research Direction

In this paper, we presented a simple, inexpensive but fully functional Context-Sensitive Cloud-based Mobile Alert and Optimal Route Discovery and Rating for ITS infrastructures. Our goal is to provide a transportation system that will allow traffic data including alerts to be aggregated into a cloud environment where such data can be analyzed by our proposed hybrid Data Analytic algorithm. Our algorithm is derived by combining enhanced features of Apache Spark and Kafka frameworks for efficient real-time analysis and alert/traffic message dissemination to ITS users. Our proposed Data Analytic Engine performs two essential functions including analysis of received alert messages to detect traffic anomalies and the analysis of traffic data to generate required Traffic Model for a particular road user based on the user profile as received from the mobile devices. We anticipated instances of system faults and failures by designing two-layer alerts/traffic data aggregators by implementing a Cloud-Based Alert Aggregator (CBAA) while still maintaining a redundant Land-Based Alert Aggregator (LBAA). The LBAA acts as a cache for offline access to traffic messages when the CBAA experiences faults or failures.

In our future research, we plan to implement, investigate and analyze the performance of the proposed algorithms operating within the Cloud-Based ITS architecture as the algorithms support real-time and high quality communication between road users mobile devices and the CBAA. We will also investigate the validity of the traffic parameters used in this research along with their assigned weights, so as to establish their correctness. We anticipate that there will be some other design and implementation issues that will emanate as a result of the proposed architecture. Therefore, we will address these challenges and related issues in our future publications.

References

1. Toroyan, T.: Global status report on road safety. Inj. Prev. **15**(4), 286–289 (2009)
2. Zheng, K., Zheng, Q., Chatzimisios, P., Xiang, W., Zhou, Y.: Heterogeneous vehicular networking: a survey on architecture, challenges, and solutions. IEEE Commun. Surv. Tutor. **17**(4), 2377–2396 (2015)
3. U.S. Department of Transportation (2020) About ITS - Frequently Asked Questions. https://www.its.dot.gov/about/faqs.htm. Accessed 28 Apr 2020
4. U.S. Department of Transportation Effects on Intelligent Transportation Systems Planning and Deployment in a Connected Vehicle Environment (2018). https://ops.fhwa.dot.gov/publications/fhwahop18014/fhwahop18014.pdf. Accessed 24 Apr 2020
5. Ezell, S.: Explaining international IT application leadership: intelligent transportation systems. ITIF-The Information Technology & Innovation Foundation, Washington, DC (2010). https://itif.org/files/2010-1-27-ITS-Leadership.pdf. Accessed 3 May 2020
6. Vanajakshi, L., Ramadurai, G., Anand, A.: Intelligent Transportation Systems Synthesis Report on ITS Including Issues and Challenges in India, Centre of Excellence in Urban Transport (2010)
7. Young, K.: Overcoming Barriers to ITS Deployment in Korea, Presentation to the ITS World Congress (2008)
8. Li, R., Jiang, C., Zhu, F., Chen, X.: Traffic flow data forecasting based on interval type-2 fuzzy sets theory. IEEE/CAA J. Autom. Sinica **3**(2), 141–148 (2016)
9. Vorhies, B.: The Big Deal About Big Data: What's Inside-Structured, Unstructured, and Semi-Structured Data (2013). http://data-magnum.com/the-big-deal-about-big-data-whats-inside-structured-unstructured-and-semi-structureddata/. Accessed 24 Apr 2020
10. Brooks, R.R., Sander, S., Deng, J., Taiber, J.: Automobile security concerns. IEEE Veh. Technol. **4**(2), 52–64 (2009)
11. Zhu, L., Yu, F.R., Wang, Y., Ning, B., Tang, T.: Big data analytics in intelligent transportation systems: a survey. IEEE Trans. Intell. Transp. Syst. **20**(1), 383–398 (2019). https://doi.org/10.1109/TITS.2018.2815678

12. Hilbert, M., López, P.: The world's technological capacity to store, communicate, and compute information. Science **332**(6025), 60–65 (2011)
13. Wang, W., Krishnan, R., Diehl, A.: Advances and Challenges in Intelligent Transportation: The Evolution of ICT to Address Transport Challenges in Developing Countries (2015). https://openknowledge.worldbank.org/handle/10986/25006. Accessed 24 Apr 2020
14. Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A.S., Buyya, R.: Big Data computing and clouds: trends and future directions. J. Parallel Distrib. Comput. **79**(80) 3–15 (2013)
15. Waze Apps (2020). https://www.waze.com/. Accessed 3 May 2020
16. Chowdhury, M., Apon, A., Dey, K.: Data Analytics for Intelligent Transportation Systems. Amsterdam, The Netherlands (2017)
17. Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., Shook, E.: Mapping the global Twitter heartbeat: the geography of Twitter. **18**(5)(2013). https://firstmonday.org/article/view/4366/3654. Accessed 3 May 2020
18. Inrix Traffic Apps (2016). http://inrix.com/mobile-apps/. Accessed 3 May 2020
19. Moovit Apps (2018). http://moovitapp.com/. Accessed 3 May 2020
20. Gaode Apps2 (2019). http://gaode.com/. Accessed 3 May 2020
21. Balogun, V.F., Obe, O.O., Balogun, T.M.: Location-based mobile alert system for intelligent transportation system. Int. J. Adv. Res. Eng. Appl. Sci. (IJAREAS) **3**(1), 11–26 (2017)
22. Ringhand, M.: Factors influencing drivers' urban route choice. Ph.D. Dissertation (2019). https://www.researchgate.net/publication/335422150-Factors-influencing-drivers%27-urban-route-choice. Accessed 24 Apr 2020
23. Bovy, P.H.L., Stern, E.: Route Choice: Wayfinding in Transport Networks, vol. 9. Springer, Dordrecht (1990)
24. Segadilha, A.B.P., da Penha Sanches, S.: Identification of factors that influence cyclistś route choice. Procedia-Soc. Behav. Sci. **160**, 372–380 (2014)
25. Tawfik, A.M., Rakha, H.A.: Network route-choice evolution in a real-time experiment: a necessary shift from network to driver oriented modeling. In: 91st Annual Meeting of Transportation Research Board Paper Compendium DVD 12-1640 (2012)
26. Papinski, D., Scott, D.M., Doherty, S.T.: Exploring the route choice decision-making process: a comparison of planned and observed routes obtained using person-based GPS. Transp. Res. Part F: Traffic Psychol. Behav. **12**(4), 347–358 (2009). https://doi.org/10.1016/j.trf.2009.04.001
27. Chamali, H., Baman Bandara, W.S.: Analysis of factors affecting pedestrian route choice. J. Chem. Inf. Model. **53**(9), 1689–1699 (2013)
28. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numerische mathematik **1**(1) 269–271(1959)
29. Farooq, M.U., Shakoor, A., Siddique, A.: An efficient dynamic round robin algorithm for CPU scheduling. In: IEEE International Conference on Communication, Computing and Digital Systems, pp. 244–248 (2017)
30. Seshasayee, A., Lakshmi, J.V.N.: An insight into tree based machine learning techniques for big data analytics using Apache Spark. In: International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), pp. 1740–1743 (2018)
31. Sarumi, O.A., Leung, C.K., Adetunmbi, O.A.: Spark-based data analytics of sequence motifs in large omics data. Procedia Comput. Sci. **126**, 596–605 (2018)

32. Le Noac'h, P., Costan, A., Bougé, L.: A performance evaluation of Apache Kafka in support of big data streaming applications. In: IEEE International Conference on Big Data (Big Data), pp. 4803–4806 (2017)
33. Jiang, F., Leung, C.K., Sarumi, O.A., Zhang, C.Y.: Mining sequential patterns from uncertain big DNA in the Spark framework. In: IEEE BIBM, pp. 874–88 (2016)
34. Matei, Z., et al.: Apache Spark: a unified engine for big data processing. Commun. ACM **59**(11), 56–65 (2016)
35. Sarumi, O.A., Leung, C.K.: Scalable data science and machine learning algorithm for gene prediction. In: The 7th International Conference on Big Data Applications, pp. 118–126 (2019)
36. Cardoso, P.V., Barcelos, P.P.: Definition of an architecture for dynamic and automatic checkpoints on Apache Spark. In: IEEE 37th Symposium on Reliable Distributed Systems (SRDS), pp. 271–272 (2018)
37. Venil, P., Vinodhini, G., Suban, R.: Performance evaluation of ensemble based collaborative filtering recommender system. In: IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1–5 (2019)
38. Thepade, S.D., Kalbhor, M.M.: Ensemble of machine learning classifiers for improved image category prediction using fractional coefficients of Hartley and sine transforms. In: Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1–5 (2018)